Machine Learning for Bank Failure Prediction: Enhancing Early Warning Systems

Khalid Mohammed, Coleman Pagac, Rediet Ayalew, Braedon Stapelman
(Supervised by Eric Manley and Sean Severe)

Drake UNIVERSITY

Introduction

Ensuring financial system stability requires early identification of institutional distress, particularly within the banking sector. Traditional statistical models face significant challenges due to sparse failure events and highly non-linear financial indicators. This research proposes a hybrid machine learning framework that improves early warning accuracy by combining classification and regression techniques with both shallow and deep learning models. Our approach captures underlying temporal patterns and complex interactions between variables indicative of bank health deterioration, enabling proactive, data-driven regulatory intervention before crisis points emerge.

HEALTH Score

We construct a composite financial metric, **HEALTH** (Holistic Evaluation of Asset Liquidity, Transparency, and Holdover stability), which is defined as (Equity - Goodwill) / Total Assets. This ratio serves as a comprehensive proxy for financial well-being, capturing core dimensions of banking performance: capital adequacy, asset quality, liquidity, and profitability. Following Wheelock and Wilson (2000), banks falling below a HEALTH threshold of 0.02 are flagged as at-risk. Unlike binary failure labels, this continuous target enables both granular risk assessments and earlier detection of deterioration trends. HEALTH is employed in both classification (thresholded) and regression (direct prediction) contexts, allowing for rich interpretability and dynamic modeling flexibility across varying regulatory needs.

Feature Engineering

Preprocessing begins with z-score normalization and log-transforming skewed indicators to ensure distributional stability. We emphasize lower HEALTH ranges via exponential rescaling, improving model sensitivity to early signs of deterioration. Class imbalance, critical in the rare-event nature of bank failures, is addressed with weighted loss functions and SMOTE augmentation techniques. We use a fixed four-quarter input window (specifically quarters 6-9 prior to the prediction quarter), allowing our models to make predictions six quarters in advance using a full year of historical financial data. This approach captures essential financial context while providing regulators sufficient lead time for intervention. Recursive feature elimination combined with SHAP value analysis helps select high-impact, interpretable features without introducing overfitting or collinearity, resulting in a final set of 24 key indicators.

Model Architecture

Our modeling suite includes logistic regression, random forests, and XGBoost for benchmark performance and interpretability. We then explore temporal deep learning models: bidirectional LSTMs and GRUs capture sequential patterns in financial data, CNN-LSTM hybrids extract both local features and temporal dependencies, and transformer architectures leverage attention mechanisms for tracking long-range dependencies across multiple time windows. All models are trained on financial time series data spanning 2003-2020, with optimized hyperparameters, dropout regularization, and careful treatment of class imbalance through stratified sampling. Cross-validation is performed on rolling time windows to maintain temporal integrity and prevent data leakage across economic cycles.

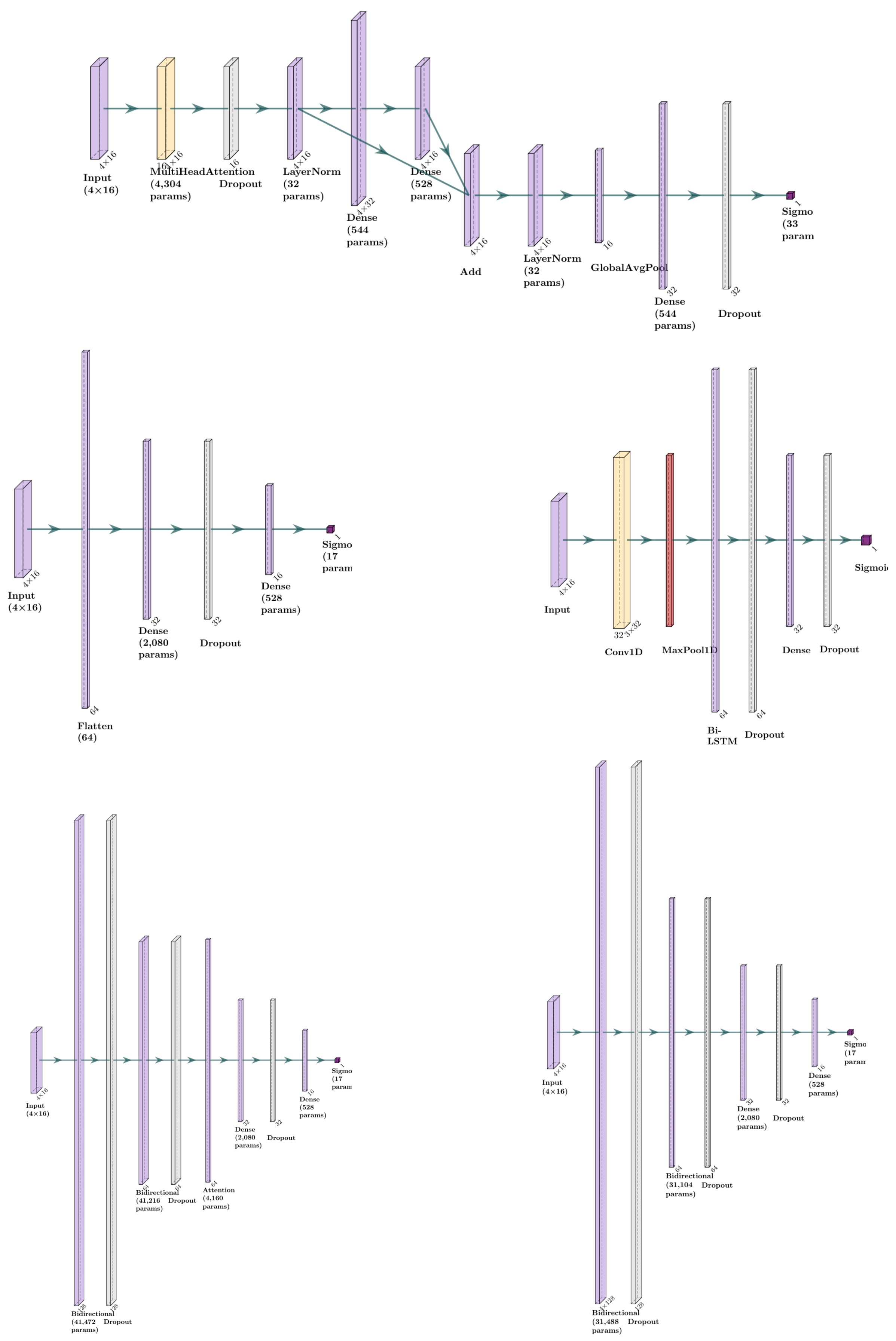

Figure 1: Architectures — Transformer, Feedforward, CNN-LSTM, BiLSTM + Attention, and BiLSTM

Results

XGBoost consistently delivers the highest AUC (0.93) for classification tasks. Among deep learning models, Transformers (AUC 0.892), Attention LSTM (AUC 0.891), and Dual-Recurrent networks (AUC 0.889) lead performance. Logistic Regression (AUC 0.78) and Random Forest (AUC 0.71) show more modest results. Feature importance analyses highlight liquidity ratios (SHAP 0.23) and capital adequacy metrics (SHAP 0.19) as key predictors. These models enable regulatory intervention up to six quarters before traditional warning signs appear.

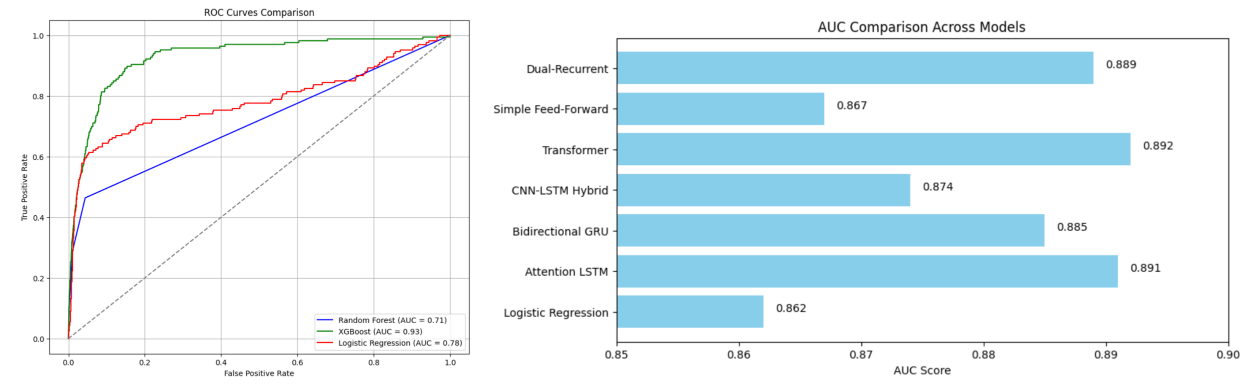

Figure 2: Left: ROC curves for traditional models. Right: AUC comparison across deep learning architectures

Conclusion

Our findings validate ML architectures as valuable tools for bank failure prediction. The HEALTH framework supports both risk flagging and continuous monitoring. XGBoost delivers superior classification (AUC 0.93) while transformer architectures (AUC 0.892) excel at capturing deterioration patterns. Liquidity ratios and capital adequacy emerge as critical predictors. ML-enhanced systems can identify at-risk institutions up to six quarters before traditional warning signs, providing regulators crucial lead time for intervention.

References

References

[1] Federal Deposit Insurance Corporation (FDIC). (2019). *Bank Financial Health Metrics: A Comprehensive Rating System.* FDIC Quarterly, 13(2), 37-51.

[2] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need.* Advances in Neural Information Processing Systems (NeurIPS), 30, 5998-6008.

[4] Wheelock, D. C., & Wilson, P. W. (2000). *Why do banks disappear? The determinants of US bank failures and acquisitions.* Review of Economics and Statistics, 82(1), 127–138.